

Juliane Stiller, Maria Gäde and Vivien Petras, Berlin

Multilingual Access to Digital Libraries: The Europeana Use Case

The article determines multilingual access features in digital libraries with a special focus on cultural heritage digital libraries. An analysis of existing information systems in the GLAM-domain (galleries, libraries, archives and museums) was conducted to establish and collect solutions for searching, browsing and interacting with multilingual content. In particular, Europeana, the European digital library, archive and museum for cultural heritage was studied with a focus on multilingual interactions. Challenges and recommendations for the implementation of multilingual access features are presented and discussed.

Keywords: multilingual search, digital library, cultural heritage, Europeana

Mehrsprachiger Zugang zu Digitalen Bibliotheken:

Europeana

Der Artikel fasst Komponenten für einen mehrsprachigen Zugang in digitalen Bibliotheken zusammen. Dabei wird der Fokus auf Bibliotheken für das digitale Kulturerbe gelegt. Eine Analyse aktueller (existierender) Informationssysteme im sogenannten GLAM-Bereich (Galerien, Bibliotheken, Archive, Museen) beschreibt angewandte Lösungen für die Recherche (Suchen und Blättern) von und die Interaktion mit mehrsprachigen Inhalten. Europeana, die europäische digitale Bibliothek für Kulturerbe, wird als Fallbeispiel hervorgehoben und es werden beispielhaft Interaktionsszenarios für die mehrsprachige Recherche vorgestellt. Die Herausforderungen in der Implementierung von Komponenten für den mehrsprachigen Informationszugang sowie Empfehlungen für den verbesserten Einsatz werden vorgestellt und diskutiert.

Deskriptoren: Digitale Bibliothek, Kulturerbe, Mehrsprachigkeit, Recherche, Europeana

Accès multilingue aux bibliothèques numériques:

Le cas d'Europeana

L'article détermine les caractéristiques d'accès multilingues dans les bibliothèques numériques et en particulier

dans les bibliothèques numériques dédiées au patrimoine culturel. Une analyse de systèmes d'information existants dans le domaine GLAM (galeries, bibliothèques, archives, musées) a été menée à bien pour fixer et percevoir des solutions pour la recherche et la navigation de contenu multilingue. En particulier, Europeana, la bibliothèque numérique européenne dédiée au patrimoine culturel, a été étudiée avec un accent sur les interactions de recherche multilingue. Les défis de la mise en œuvre des fonctions d'accès multilingues et des recommandations sur la manière de les surmonter sont ici présentés et discutés.

Mots-clés: recherche multilingue, bibliothèque numérique, Europeana, patrimoine culturel

1 Introduction

Multilingual access to digital content is of particular interest for political, legal and cultural entities that deal with documents in different languages and have audiences with diverse language backgrounds. Bridging the language gap is essential for most European organizations and helps to create a common European identity based on shared values and cultural similarities. This is especially true for digital libraries providing access to our cultural heritage; they need to find strategies to encourage the discovery of content in languages one might not understand. To overcome these language barriers, solutions from different areas come into play. On the one hand, these solutions address system-centred problems ranging from automatic translation, data management, character representation, software development and interoperability issues (Diekema, 2012). On the other hand, there are user-centred solutions that deal with multilingual user interaction and interface design. This paper provides an overview of the different dimensions of multilinguality in cultural heritage digital libraries. It examines the most common user interactions with respect to multilinguality: searching, browsing and engaging. The analysis of several digital libraries in the cultural heritage domain resulted in an overview of implemented multilingual features and the challenges that arise when

displaying content in different languages to users. Europeana¹, the multilingual digital library providing access to digital representations coming from museums, archives, galleries and libraries, serves as a special use case. Its multilingual access points and different features for exploring content across languages are discussed in detail. The goal is to derive recommendations for digital libraries for multilingual solutions to provide and simplify access across languages².

2 Dimensions of Multilinguality in Digital Libraries

Language diversity forms a wall hindering users from accessing and exploring content in the web and digital libraries (Large and Moukdad, 2000). To overcome this, digital libraries need to ensure that users can convey meaning from content that is not presented/available in their native or preferred language. Four main levels of multilingual access in information systems can be distinguished:

- multilingual display
- multilingual search and browsing
- multilingual result representation and translation
- multilingual user engagement

An important distinction needs to be made between the offering of

- multilingual interface options (static web pages)
- multilingual objects such as full text books in different languages
- or multilingual metadata.

Peters et al. list four support tools, which enable cross-language retrieval and browsing and multilingual result presentation: query formulation support, evaluation support for the selection of documents, support for query reformulation and means for browsing collections and results (Peters et al., 2012, p. 96). Usability and the design of the interaction in these tools is a concern especially as the interface gets cluttered with different language options leading to potential confusion on the user side.

¹ <http://www.europeana.eu/>

² The article is based on a deliverable the authors wrote for the EU-funded project Europeana v2.0 “Midterm Report on Innovative Multilingual Information Access” (Stiller et al., 2012). It gives an account of the research conducted during the first half of the project. The deliverable is available online: <http://pro.europeana.eu/documents/866067/b0103ac0-611f-4a04-b4b6-f28e200a04e1>

2.1 Multilingual Display

The localization or internationalization of interfaces is the basic level of multilingual information access. The customization of the interface according to the user's preferred or native language assures that users can understand and navigate a website irrespectively of their native language. Automatic interface language changes are realized through cookies that store information about the users' origin or preference based on the language of the browser or the country the user is coming from using his geo-location information. User-assisted language changes can be either provided via drop-down menus, buttons or flags. While the localization or internationalization of the interface language seems to ensure multilingual access, this is only true for mostly static web page content (for which internationalization was applied). The digital content, for which the information system was developed, commonly resides behind a database or search engine index and is not affected by this step. In order to provide multilingual search and browsing options, all digital content or metadata needs to be translated.

2.2 Multilingual Search and Browsing

Multilingual searching is usually referred to as cross-language information retrieval (CLIR). CLIR allows users to find documents or their metadata in several languages irrespectively of the query language, meaning that the language of the query (a user's language) and the language of the content do not need to be the same (Oard and Diekema, 1998). To overcome the language barrier between the query language and the object language, different solutions have been suggested (Oard and Diekema, 1998, Oard, 1998). The most common approach is query translation, where the user queries are translated into the content language(s). The query translation process requires several steps, which are either performed automatically or with assistance from the user. The first step is the identification of the query language, which is a cumbersome task due to their short length. Even for human assessors the language of a query is hard to determine as ambiguities across languages increase with each language added to the pool (Stiller et al, 2010). Some systems require the user to determine the query language and target language whereas other systems perform hidden multilingual search. Once the query language is identified, the query needs to be processed and translated into the content language(s). If the collection contains documents in different languages, the query

needs to be translated into all available content languages. Systems that support user-assisted query translation usually present translation candidates from which the user can choose the most appropriate ones.

Another approach for multilingual search is the translation of documents into all user languages a system wants to support. Research showed downsides to document translation such as time and storage consumption and maintaining translations over time (Peters et al., 2012, p. 59). This especially applies to systems supporting many different languages. A compromise is using a pivot language into which both queries and documents are all translated so the search problem becomes essentially monolingual.

Multilingual search and browsing can also be facilitated through the use of multilingual vocabulary incorporated to enrich the describing metadata. This allows the users to retrieve items across languages, find related objects and browse items by concepts or geographic location depending on which metadata fields got enriched. Multilingual browsing helps users to access content and discover objects they were not aware of. Especially if users do not speak the languages the objects or their metadata are in, it is essential to offer browsing capabilities that support serendipity and discovery of the unknown. Support of multilingual browsing can also be accomplished by providing facets for refining results in specific languages or offering curated guided tours in several languages.

2.3 Multilingual Result Representation and Translation

The representation or even translation of retrieved objects from different countries in several languages poses another challenge for multilingual information systems. Systems can either display all results in one merged list or separated by language. Users should be able to refine search results according to their preferred language(s) and be able to have results translated they cannot understand. System designers need to make the choice when and to which degree users should be confronted with decisions on how to handle multilingual results. One example would be that the users indicate into which language the results should be translated during the query input or after the results are retrieved.

In search result representation, support tools need to be designed to help users to determine the relevance of objects that are provided in foreign languages. It can be helpful to visually emphasize translation alignments

helping users to find reformulations for their queries and assessing the relevance of the retrieved objects.

On the object or document level, it needs to be determined if a full translation is desired or if metadata translation is sufficient. If objects are available in several languages, the display of the multilingual data is a concern. The main questions here are:

- Which criterion is used to determine the display language?
- How can the user switch between different language versions of objects?

To effectively support multilingual users in retrieving information in languages they might not understand, information systems need to offer means to help users through the information seeking process.

2.4 Leveraging Collaboration and Engagement for Multilingual Access

Collaboration and engagement features such as social tagging and user-curated exhibitions are more and more included in the feature set of cultural heritage digital libraries. The goal is to complement the visit to a cultural institution and create a space where the interaction with the cultural items can be personalized. With these developments, the possibility arises to leverage the user input into the system to derive translations and improve existing dictionaries. One example is social tagging features. In an information system with users from different linguistic backgrounds the potential amount of social tags in different languages is quite high. Aligning translation pairs can be beneficial for improving multilingual access. A study from Eleta and Goldbeck (2012) on social tags' potential to bridge language gaps concluded that power tags in different languages happen to be translations of each other. Nevertheless, cultural differences have an influence on the choice of tags and especially historic references contrast in different cultural environments (Eleta and Golbeck, 2012). Similarly, queries in different languages can be mapped and used for improving translation resources. Based on the assumption that users type the same query in several languages in a multilingual information system, an automatic approach to aggregate these translation pairs from query logs was developed (Bosca and Dini, 2009).

Table 1. Type of sites in the sample of cultural heritage digital libraries.

| Site type | Definition | Example |
|-------------------|---|--|
| Museum (11) | Websites providing access to the resources of a museum and visitor information | www.louvre.fr |
| Archive (3) | Websites providing access to the resources of an archive and visitor information | www.nationaalarchief.nl |
| Library (3) | Websites providing access to the resources of a library and visitor information | www.perseus.tufts.edu/hopper |
| Aggregator (7) | Websites offering a single access point to the resources of several institutions or organizations. Here the affiliation to a certain region or type of organization is the main characteristic. | www.europeana.eu |
| Collection (4) | Websites offering a single access point to resources that are united by a theme retrievable in the content. | www.philaplace.org |
| Community (3) | These websites are living from and for the content of the user and the community. They can be arranged around a theme or a specific topic. | www.historypin.com |

3 Multilingual Information Access in Cultural Heritage Digital Libraries

In order to provide a more structured overview of applications of multilingual interaction features, an analysis of 31 cultural heritage digital libraries was conducted. Table 1 shows the types of sites examined in this analysis. The definitions of museum, archive, library and community sites were developed on the basis of a survey initiated by OCLC on social metadata (Smith-Yoshimura and Shein, 2011); the other categories were developed based on the requirements of this study.

Roughly half of the analysed websites originate in Europe; the rest is located in the U.S. The majority deals with content in different languages, as cultural heritage collections are often linguistically diverse. The most common observed multilingual feature is the interface language change. Almost all sites from Europe support such a feature but only a couple of sites coming from the U.S. Some of the websites used third party solutions for the interface translation such as Google Translate³, others relied on in-house translation. The describing metadata is mainly monolingual depending on the hosting institution and the anticipated audience for the content, i.e. translation needs to be applied in order to provide multilingual access to content. Nevertheless, to gain greater reach, websites do offer their metadata in one more language, mainly English, if it is not the primary language. Monolingual metadata is prevailing in smaller projects and orga-

nizations originating in the United States. If the institution is known across borders and has international significance such as the Google Art project⁴, metadata is more likely to be in several languages. In some cases, the metadata language depends on the language the digital object was provided in, as it is the case for Europeana. On some websites, the metadata language also changes with the interface language chosen by the user, but in most cases, the interface language change does not influence the language of the metadata. One website, the International Children's Digital Library⁵, offers a particularly rich multilingual experience with the goal to make its content accessible across countries. They acknowledge the fact that users with specific language backgrounds have specific and varying user needs (Bilal and Bachir, 2007).

The following sections discuss different multilingual features offered by the analysed cultural heritage digital libraries in more detail.

3.1 Multilingual Display in Cultural Heritage Digital Libraries

Through localization, the content of all static interface pages can be translated into the users' preferred language. In most surveyed cases, this does not include the translation of metadata. If institutions offer metadata translation, then it is often through automatic means translating the content on the fly when the user is requesting it. This can lead to clashes in the display, as the

³ <http://translate.google.com/>

⁴ <http://www.googleartproject.com/>

⁵ <http://en.childrenslibrary.org>

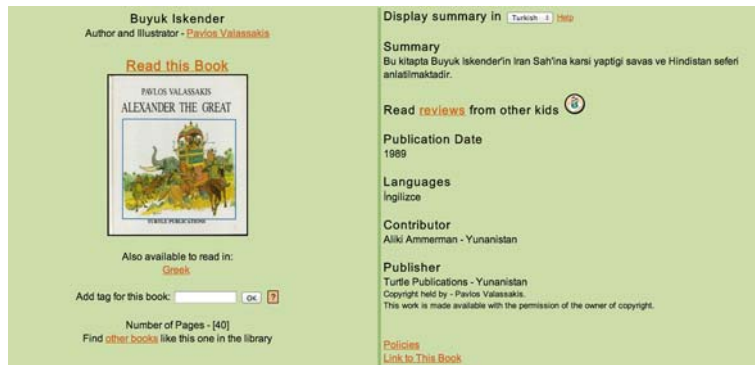


Fig. 1: ICDL – English book with metadata translation in Turkish.

language of the navigation might be different than the language of the metadata.

The challenge of providing all the content in different languages is augmented by the problems that arise when presenting different language versions to the user – problems of visual design, usability and interaction design need to be tackled. A good example of a website presenting multilingual content is The International Children's Digital Library (ICDL). Figure 1 gives an example of an object that is originally published in English. The whole book is digitized and can be read on screen in English and in Greek added by volunteer translators. The display summary has its own language drop-down menu that translates the field values into the preferred languages, in this case Turkish.

To display multilingual content, different solutions were implemented in the sample websites:

- multilingual display of static content
- multilingual display of dynamic content
- several multilingual versions of the content and the website: a certain language version is chosen as entry point and you can not switch between versions (e.g. State Hermitage Museum⁶)
- parts of the websites and the digital objects are translated separately.

3.2 Multilingual Search and Browsing in Cultural Heritage Digital Libraries

Multilingual information retrieval is rarely implemented in the cultural heritage domain although it provides the ideal use case as cultural heritage collections are often multilingual and audiences come from different countries. Technical obstacles might hinder developments, as

cross-lingual information retrieval is requiring query or object/metadata translation. None of the websites had query translation implemented. Nevertheless, more and more sites enable search in different languages through multilingual controlled vocabularies or metadata in different languages. Especially, in fielded location search, vocabularies such as GeoNames⁷ are used to enable cross-lingual retrieval. Another approach is to use the Google Maps API that comes with multilingual versions of geographic names.

Multilingual browsing was rarely implemented in the sample cultural heritage websites. Nevertheless, more and more browsing features are based on spatial and time information in the metadata and objects are then displayed in a non-textual way, which can be often understood across languages. Many websites use APIs of map providers to plot their data on maps and let users discover content based on a geographic location. Map and timeline browsing allow viewing the collection from a different perspective and might support a language-independent (or multilingual) discovery process.

3.3 Collaborative Features

Only a few websites in the cultural heritage domain offer collaborative features for users. Reasons for this are manifold, on the one side there is the fear that user input might result in poor quality content; on the other side there is the technical complexity of such a system. Another pitfall is to set the right incentives for users to participate. Many websites in the cultural heritage domain offer great features but are missing a solid user base. The potential of collaborative features to enrich and contextualize existing content should not be underestimated.

⁶ <http://www.hermitagemuseum.org/>

⁷ <http://www.geonames.org/>

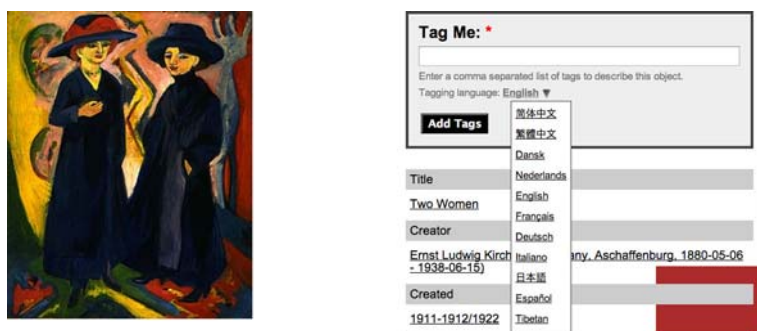


Fig. 2: Steve Tagger Project – drop-down menu to specify the language of tags.

They can also help to bridge language gaps and connect digital objects through users adding additional access points.

In the sample, the most common implemented feature is social tagging. However, most of these social tagging features are not designed to aggregate multilingual tags or annotations; they are targeted for monolingual use. A strategic implementation of social tagging that guides the user through the workflow ensures high quality of the tags. Dictionaries and controlled vocabularies reduce the risk of misspellings and help the user to disambiguate homonyms and named entities. If the user is asked to provide the language of his tags, they can be used to enrich digital objects multilingually. An example of such a system is the Steve Tagger Project⁸, which prompts the user to specify the language of their assigned tags (Figure 2).

Active calls to participate in enriching metadata and helping adding more information to digital objects are rare and in most cases they are dealing with monolingual content. Collaborative translation features require complex systems to manage the user input and to ensure quality. Cultural heritage websites manage translation loads they want to outsource to the public by recruiting volunteers. In this case, the community does not oversee the quality and process of the translation; this is rather done by individuals who are managed by the offering institution. For example, the ICDL recruits volunteers who translate the books, metadata and the web interface into other languages⁹. Quality control and the responsibility for correcting the translations stay with the institution. It is obvious that this solution requires a high level of maintenance and the process is not self-sustainable. On the other side, handing over the translation aspect completely to the community requires a community strategy,

an engaging and usable web interface and community managers who enforce rules and guide the process.

4 Multilinguality in Europeana

Europeana is an interesting use case for analysing multilingual information access as its over 25 million digital objects (as of March 2013) are in many different languages serving a user base with diverse language backgrounds. It is Europeana's goal to push for new and innovative multilingual access strategies and act as a trailblazer in the cultural heritage domain. Several surveys focusing on multilingual access to Europeana were conducted (Agosti et al., 2009; IRN Research 2009 & 2011).

On average, respondents have language skills in at least 1.5 other languages and 71% of all non-English native speakers could access and interact with websites in English (IRN Research 2011, p. 9) showing the multilingual capabilities of the Europeana users. In general, users feel comfortable accessing the portal and scanning results in their native language or in English. A “significant language barrier was perceived” when users had to deal with content in unknown languages (Dobrev and Chowdhury, 2010). The most popular result refinement options in Europeana are the language and country facets providing an indicator for the importance of language and geographic search options (IRN Research 2009, p. 3). Dobrev and Chowdhury (2010) also found out that a strong need for more content in native languages as well as result translation options exists. The majority of users (80%) are willing to control the query translation process meaning that active user control for multilingual access would be acceptable (Agosti et al., 2009). No clear preference was found regarding the multilingual result representation.

The following sections give an overview of multilingual access features in Europeana and present the strategies, which proved to be successful in letting users explore multilingual content they are unaware of.

⁸ <http://tagger.steve.museum/>

⁹ <http://en.childrenslibrary.org/contribute/translate.shtml>

4.1 Multilingual Display in Europeana

Europeana offers its static content in 31 different European languages. By default, first time users get the English interface and they can switch to their desired language by choosing it from a drop-down menu. Once the user switched to his preferred language, a cookie is set which directs him to his selected interface on the next visit. For future visits, Europeana is delivered in the chosen language as long as cookies are enabled and not deleted in the meantime.

4.2 Multilingual Search

To enable search across languages and support the discovery of related items, Europeana enriches its content with multilingual vocabularies. Certain metadata fields for subject, date, place and creator, are enriched with a specific vocabulary. For example, the metadata fields for `dc:subject` and `dc:type`, which generally specify concepts, were enriched with the GEMET Thesaurus¹⁰, a multilingual environmental thesaurus.

Enriching subject metadata fields with all translation equivalents of a concept enables the user to retrieve documents, which are not written in the language of his query. One example is the query “cheval” (fr. horse) retrieving the Russian object “Лошадь, конюх и собака.”, which has the Russian term for horse in its title ‘Лошадь’. The object was enriched with translations for “horse” from the GEMET thesaurus and could therefore be retrieved by a query in another language. In Europeana, the associated tags can be found in the right side bar under ‘Auto-generated tags’.

However, in a qualitative study conducted at the Berlin School of Library and Information Science it was shown that enrichments could introduce errors when not implemented with an enrichment strategy (Olensky et al., 2012).

4.3 Multilingual Browsing

Multilingual browsing capabilities are essential for Europeana to enable its users to understand extent and scope of the collections and support serendipity and exploration of unknown cultural material in languages, which are not understood by the user.

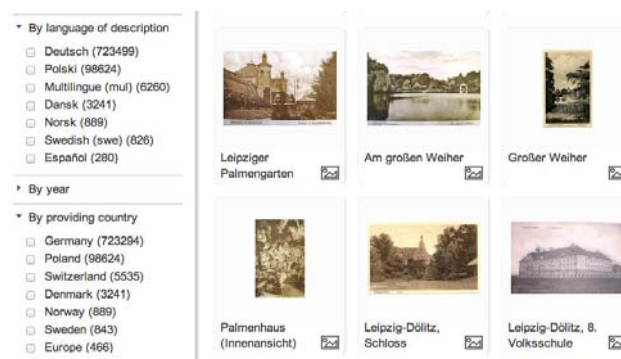


Fig. 3: Facets for refining search results related to the language of the description and the country of origin (<http://preview.europeana.eu/>).

Europeana offers two language-agnostic features enabling users to see results for their queries on a timeline or a map. Theoretically, this could enable browsing across languages but the features are only special result displays requiring a query input. Results are then mapped to the geographic location the object is provided from and mapped on a timeline according to the date specified in the metadata.

Europeana offers several facets to filter retrieved results. The language facet allows users to refine objects according to the metadata language (Figure 3). This language might not correlate with the actual language of the object as it indicates the language of the describing metadata. Problems like this show how complex multilingual information is and that there might be a gap between the user’s expectations and the offerings of the digital library.

In addition, Europeana curates virtual exhibitions around relevant themes. Most of these exhibitions are offered in different languages (Figure 4) enabling a contextualization of the content in different languages.



Fig. 4: Europeana exhibition “Art Nouveau” offered in different languages.

¹⁰ <http://www.eionet.europa.eu/gemet/>

4.4 Object Translation in Europeana

Result or object translation is currently offered by an external translation service provided by Microsoft. Previous studies have shown that users are satisfied with metadata translation and do not require full text translation in order to assess results (Oard et al., 2004). Users can select their preferred language via a drop down menu and translate metadata information including the object description. The object title as well as the metadata field names are not translated but remain in the source or selected interface language. At any time users can go back to the original representation.

5 Recommendations for Multilingual Access in Digital Libraries

Multilingual interactions patterns or access features are still not much discussed in literature and the analysed websites do not follow any obvious conventions. The consequences of using multilingual access features are often not transparent and might be confusing for the user; for example, do users understand the consequences and limits when switching the interface language via a drop-down menu (a reappearing misunderstanding is that with the interface language switch also a content language switch occurs)? Clear labelling is desired; making sure feedback is given on features that have a multilingual dimension.

From the survey of implemented multilingual components, we can derive general recommendations for the application of multilingual features in digital libraries.

5.1 Multilingual Display

To put the display and workflows in place to navigate and retrieve content, which is different from the user's preferred language is not a trivial task. The mixing of languages should be avoided. This often happens when static pages are translated and parts of the dynamic content remain in the original language, e. g. static metadata field names are translated but the content of the fields is staying in the original language. For the user, it should be possible to switch between the different interface versions on every page of his path navigating through the website and the landing page the user might be directed to from web search engines or external links. Having se-

parate entities of the website is not recommended as it locks the user into a particular language version.

The user should be aware, which of his actions influence the multilingual appearance. It should be clear to the user which parts of a website are translated when using a language drop-down menu. Often, the language-specific features are not self-explanatory and in the worst case misleading. In Europeana, one of the language specific facets that lets users refine search results is called "language" listing different European languages. Users cannot know whether this refers to the language of the object, its metadata or the language of the providing institution. In the preview of the new Europeana version¹¹, the facet is renamed to "language of description" making it clearer, on which characteristic the facet is refining the results (see also Figure 3).

5.2 Multilingual Search

A lot of research was conducted on advancing cross-lingual information retrieval and the technology to find documents with queries whose language differs from the document. In the cultural heritage domain, research in this field is scarce, although this area deals with additional issues such as heterogeneity of the data combined with problems of metadata quality. With regards to retrieval performance, good results were achieved when flattening the metadata structure and treating cultural heritage metadata as free text (Koolen et al., 2007). Retrieval performance with multilingual cultural heritage data is now also evaluated within the CLEF (cross-lingual information retrieval) initiative CHiC¹² that focuses on the evaluation of multilingual retrieval tasks (Petras et al., 2012). The test collections are derived from the Europeana index with real life queries.

At the very least, the digital library should offer a query translation option. The best degree of interaction and user control (how automatic is the query translation) has yet to be determined.

Other features can support multilingual search by decreasing the cognitive load while helping the user to grasp the extent of the collections. For supporting the query (re)-formulation process and avoiding misspellings, autocomplete and auto-suggest features can be used (Hearst, 2009, chapter 6). Here it is essential to indicate the language the suggestion is in. Especially if the suggestions come from a multilingual controlled vocabu-

¹¹ <http://preview.europeana.eu>

¹² www.promise-noe.eu/chic-2013/home

lary, the language of the user query should match the language of the suggestion. If the user's preferred language is not known, he needs to be able to derive the language of the suggestions through the interface.

5.3 Multilingual Browsing

As with multilingual search, multilingual browsing can be implemented by enriching metadata with multilingual vocabulary. Here it is especially important to use domain-specific vocabularies, in this case from the cultural heritage domain, and to ensure that the language of the enriched term matches the language of the enriching term.

Digital libraries benefit from implementing features that let users discover content regardless of the language of its descriptions. Examples are map searches and specific characteristics like shape and colour search features. The idea is to find documents with other means than the textual representations. Nevertheless, it is important to not forget that the navigation of these features is still based on text that needs to be translated.

6 Conclusions

Multilingual information access is very important in the cultural heritage domain. The content itself is often multilingual and users who want to access cultural heritage objects come from diverse (cultural) backgrounds. Multilingual access strategies need to be implemented to allow users to bridge the language gap and support them in understanding content that is not available in their preferred language.

Europeana as the single access point to Europe's cultural heritage is unique with regards to the language diversity of its content and its multilingual audience. It acts as a trailblazer in this domain providing multilingual access and supportive interaction models to explore multilingual content and derive meaning from it. This article shows that Europeana already offers many multilingual access points. Major achievements are the multilingual enrichments of the metadata that facilitate retrieval across languages and the curated exhibitions, which highlight content in several languages. However, more work lies ahead, particularly in improving the multilingual search process.

Multilingual features can get very complex and run the risk of cluttering the user interface detracting the user from his tasks. The same is true for collaborative

features such as social tagging, from which one would like to derive translations. They require complex user management systems and strategies for maintaining and displaying tags in different languages. Strategies for such features also need to include considerations about the incentives offered for users to participate and the transparent purpose the tags will serve. To offer seamless multilingual access, a balance must be found between usability and the introduction of additional workflows that raise the cognitive efforts on the user side. Users need to be made aware of the multilingual nature of the content and the available options on how to access it with or without their preferred language. Most digital libraries have only just started to scratch the surface of possible multilingual interaction features.

References

- Agosti, M.; Crivellari, F.; Deambrosis, G.; Ferro, N.; Gäde, M.; Petras, V.; Stiller, J.: *D2.1.1 Report on User Preferences and Information Retrieval Scenarios for Multilingual Access in Europeana/EuropeanaConnect*. 2009.
- Bilal, D.; Bachir, I.: Children's interaction with cross-cultural and multilingual digital libraries. II. Information seeking, success, and affective experience. In: *Information Processing & Management*, 43 (2007), Nr. 1, S. 65–80.
- Bosca, A.; Dini, L.: The Role of Logs in Improving Cross Language Access in Digital Libraries. In: *Proceedings of the International Conference on Semantic Web and Digital Libraries*. 2009. <http://research.celi.it/attach/Publications/icsd2.pdf>.
- Diekema, A. R.: Multilinguality in the digital library: A review. In: *The Electronic Library* 30 (2012), Nr. 2, S. 165–181. doi: 10.1108/02640471211221313.
- Dobрева, M.; Chowdhury, S.: A User-Centric Evaluation of the Europeana Digital Library. In: *The Role of Digital Libraries in a Time of Global Change: Proceedings of 12th International Conference on Asia-Pacific Digital Libraries (ICADL 2010)*. Gold Coast, Australia, June 21–25 2010.
- Eleta, I.; Golbeck, J.: A study of multilingual social tagging of art images: cultural bridges and diversity. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 2012.
- Hearst, M.: *Search user interfaces*. Cambridge; New York: Cambridge University Press., 2009.
- IRN-Research: *Europeana online visitor survey – Research Report Version 3: IRN Research*. 2009.
- IRN-Research: *Europeana online visitor survey – Research Report Version 3: IRN Research*. 2011.
- Koolen, M.; Kamps, J.; Keijzer, V. D.: Information Retrieval in Cultural Heritage. In: *Interdisciplinary Science Reviews* 34 (2009), Nr. 2, S. 268–284.
- Large, A.; Moukdad, H.: Multilingual access to web resources: an overview. In: *Program: electronic library and information systems* 34 (2000), Nr. 1, S. 43–58.
- Oard, D. W.: A Comparative Study of Query and Document Translation for Cross-Language Information Retrieval. In: *AMTA '98*

Proceedings of the Third Conference of the Association for Machine Translation in the Americas on Machine Translation and the Information Soup. 1998. <http://portal.acm.org/citation.cfm?id=749060>.

- Oard, D. W.; Diekema, A. R.: Cross-Language Information Retrieval. In: M. Williams (Hrsg.): *Annual Review of Information Science and Technology* 33 (1998), S. 223–256. Medford N.J. etc.: Learned Information [etc.].
- Oard, D.; Gonzalo, J.; Sanderson, M.; López-Ostenero, F.; Wang, J.: Interactive cross-language document selections. In: *Information Retrieval* 7 (2004), Nr. 1–2, S. 205–228.
- Olenksy, M.; Stiller, J.; Dröge, E.: Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy. In: *Metadata and Semantics Research*. 2012. http://dx.doi.org/10.1007/978-3-642-35233-1_25.
- Peters, C.; Braschler, M.; Clough, P.: *Multilingual Information Retrieval: From Research to Practice*. Springer, 2012.
- Petras, V.; Ferro, N.; Gäde, M.; Isaac, A.; Kleineberg, M.; Masiero, I.; Stiller, J.: Cultural Heritage in CLEF CHiC Overview 2012 (CLEF Online Working Notes/Labs/Workshops). In: *CLEF 2012*.
- Smith-Yoshimura, K.; Shein, C.: *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews*. Dublin, Ohio: OCLC Research, 2011.
- Stiller, J.; Gäde, M.; Petras, V.: Ambiguity of Queries and the Challenges for Query Language Detection. In: CLEF 2010 LogCLEF Workshop. In: CLEF 2010 Labs and Workshops Notebook Papers. Padua, Italy, 22–23 September 2010.



Juliana Stiller, M.A.
Researcher Europeana v2.0, GALATEAS, Promise
Institut für Bibliotheks- und Informationswissenschaft (IBI)
Humboldt-Universität zu Berlin
Dorotheenstraße 26
10117 Berlin
juliane.stiller@ibi.hu-berlin.de
www.ibi.hu-berlin.de

Juliane Stiller ist wissenschaftliche Mitarbeiterin am Lehrstuhl Information Retrieval am Institut für Bibliotheks- und Informationswissenschaft. Sie arbeitet im EU-finanzierten Projekt Europeana v2.0 an neuen Interaktionsmodellen für den Zugang zu mehrsprachigen Inhalten, sowie im Projekt GALATEAS an der Evaluierung von Algorithmen zur Klassifizierung und Übersetzung von Suchanfragen. In ihrer Dissertation entwickelt und evaluiert sie Interaktionsszenarios mit kulturellen Objekten in digitalen Bibliotheken.



Maria Gäde, M.A.
Researcher Europeana v2.0 und Promise
Institut für Bibliotheks- und Informationswissenschaft (IBI)
Humboldt-Universität zu Berlin
Dorotheenstraße 26
10117 Berlin
maria.gaede@ibi.hu-berlin.de
www.ibi.hu-berlin.de

Maria Gäde ist wissenschaftliche Mitarbeiterin und Doktorandin am Lehrstuhl Information Retrieval am Institut für Bibliotheks- und Informationswissenschaft. Sie arbeitet im EU-finanzierten Projekt PROMISE (Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation) im Bereich Cultural Heritage Evaluation. Im Rahmen der CLEF Konferenz (Conference and Labs of the Evaluation Forum) ist sie an der Organisation der CHiC (Cultural Heritage in CLEF) Evaluationsinitiative, die sich mit der Evaluation des mehrsprachigen Information Retrievals in kulturellen Informationssystemen auseinandersetzt, beteiligt. Ihre Forschung bezieht sich auf die Evaluation von Digitalen Bibliotheken, im speziellen auf das Nutzerverhalten und Voraussetzungen für mehrsprachige Digitale Bibliotheken.



Prof. Vivien Petras, PhD
Institut für Bibliotheks- und Informationswissenschaft (IBI)
Humboldt-Universität zu Berlin
Dorotheenstraße 26
10117 Berlin
vivien.petras@ibi.hu-berlin.de
www.ibi.hu-berlin.de

Prof. Vivien Petras, PhD ist Professorin für Information Retrieval am Institut für Bibliotheks- und Informationswissenschaft. Seit fünf Jahren befasst sie sich in mehreren Projekten (EuropeanaConnect, Europeana v2.0) mit Aspekten des mehrsprachigen Zugangs zu Europeana. Sie ist außerdem involviert in den EU-Projekten GALATEAS (Generalized Analysis of Logs for Automatic Translation and Episodic Analysis of Searches) und PROMISE (Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation). Seit 2009 koordiniert sie die CHiC (Cultural Heritage in CLEF) Evaluationsinitiative, die sich mit der Evaluation des mehrsprachigen Information Retrievals in kulturellen Informationssystemen auseinandersetzt.